ICS 点击此处添加 ICS 号
CCS 点击此处添加 CCS 号

T/SHSOT

团 体 标 准

T/SHSOT XXXX—XXXX

动物实验标准化替代方法实施指南

Guidelines for the Implementation of Standardized Alternative Methods of Animal Experiments

草案版次选择

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX－XX－XX 发布 XXXX－XX－XX 实施

发 布

目　　次

前　　言 II
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 经验证和仿制的替代方法 5
5 可接受标准 6
6 试验设计 6
7 测试终点的室内验证 10
8 效能化学物质 12
9 数据密集型体外方法 13
附　录　A　皮肤模型刺激性测试的实施指南 15
附　录　B　（资料性）基于转染报告基因的角质细胞氧化反应活化的皮肤致敏方法的实施指南 21
参　考　文　献 25

前　　言

本文件按照GB/T 1.1-2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由上海市毒理学会提出并归口。

本文件起草单位：上海交通大学医学院，广州市华代生物科技有限公司，等

本文件主要起草人：程树军、杜娟、张超，等

本文件首批执行单位：上海交通大学医学院，广州市华代生物科技有限公司，等

动物实验标准化替代方法实施指南

1 范围

本标准适用于毒性测试动物实验的标准化替代方法的开发、验证和实施；
本标准规定了体外毒理学方法实施评价的一般性原则、程序、内容、方法和技术要求；
本标准规定的实施指南适用于化妆品、化学品及相关产品的毒性测试的替代方法；
本标准列举皮肤刺激、皮肤致敏标准化替代方法的实施作为参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

化妆品安全技术规范

OECD. 2005. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, OECD Series on Testing and Assessment No. 34, OECD, Paris.

OECD.2018. Guidance Document on Good In Vitro Method Practices (GIVIMP), OECD Series on Testing and Assessment, No. 286, OECD Publishing, Paris.

OECD.2018.Guiding Principles and Key Elements for Establishing a Weight of Evidence for Chemical Assessment Series on Testing and Assessment No. 311, OECD Publishing, Paris.

OECD 492 重建人角膜样上皮(RhCE)测试方法（Test method for reconstructed human corneal epithelium (RHCE)）

OECD 442D 体外皮肤敏感试验：针对皮肤敏化不良结果通路上角质细胞活化的关键事件检测（In Vitro Skin Sensitisation Assays addressing the AOP key event on keratinocyte activation）

OECD 430 体外皮肤腐蚀：经皮电阻测试方法（TER））（In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test Method (TER)）

OECD 431 体外皮肤腐蚀:重建人表皮(RhE)试验方法（In Vitro Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method ）

OECD 439 In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method（体外皮肤刺激:重建人体表皮试验方法）

OECD 492 重建人角膜样上皮(RhCE)测试方法（Test method for reconstructed human corneal epithelium (RHCE)）

OECD. Series on Testing and Assessment No. 220.Performance standards for the assessment of proposed similar modified in vitro reconstructed human epidermis (RhE) test methods for skin irritation testing as described in TG 439.

3 术语和定义

下列术语和定义适用于本文件。

3.1

可接受标准 Acceptance criteria

研究结果可以被接受的标准，即描述体外方法特征的一组明确定义的参数，如阳性对照和阴性对照的可接受范围等。

3.2

准确性Accuracy

指检测值与标准值或已知值的接近程度。

3.3

有害结局通路Adverse Outcome Pathway (AOP)

一种分析性结构，描述在不同的生物组织水平上的因果关联事件的连续链，这些事件导致有害的人类健康或环境效应。

3.4

基准剂量（BenchMark Dose BMD）或基准浓度（Concentration BMC）

与之前特定的生物学反应相关联的剂量或浓度，开发基准剂量/浓度应用于替代未观察到有害效应水平（NOAEL）和最低可观察到有害效应水平（LOAEL）。

3.5

最佳规范Best practice

普遍接受的最佳方法或技术，可以作为一个基准；与其他方法取得的结果相比，在不同实验室中始终显示出优异的结果。还用于描述开发和遵循被多个组织认同的标准化的实施方式及过程。最佳规范只是一个简要说明，可能会根据正在进行的科学进步而发生变化。

3.6

实验室间评估Between-laboratory assessment:

方法验证阶段，通常也称为实验室之间（或相互）验证，不同实验室的不同操作者独立执行（或运行)体外方法，以确定体外方法是否可以在不同的实验室成功建立。例如，评估实验室间再现性（between laboratory reproducibility，BLR）。

3.7

生物学通路：Biological pathway:

细胞内分子间的一系列作用，导致细胞内产生某种产物或发生变化。这种通路可以触发新分子的合成，例如脂质或蛋白质。通路还可以打开和关闭基因，或刺激细胞移动。代谢通路和信号通路是一些典型的生物学通路类型。

3.8

变异系数Coefficient of Variation (CV)

用于描述相对于均数的变异量的分散程度的参数。由于变异系数定义为与均数大小无关，且无单位，其可代替标准差，用于不同单位或不同均数数据集分散程度的比较。

3.9

细胞毒性 Cytotoxicity

一般细胞毒性(或基础细胞毒性)是对机体所有细胞共有的结构和功能产生毒性作用的结果，如 DNA、染色体、线粒体、细胞骨架和各种膜。

3.10

衍生数据derived data

衍生数据取决于原始数据，并且可以从原始数据重建，例如，通过依赖原始数据的电子表格计算的最终浓度、由 LIMS 汇总的结果表等。

3.11

数据（原始数据）Data (raw data)

数据（原始数据）可以定义为实物、过程或事件的可测量或描述性的属性。GLP 原则将原始数据定义为所有实验室记录和文件，包括通过自动仪器界面直接输入计算机的数据，这些数据是一项研究中主要观察和活动的结果，是重建和评估研究报告所必需的。

3.12

数据的完整性Data integrity

所有数据在整个数据生命周期中完整、一致和准确的程度。

3.13

数据生命周期Data lifecycle

数据（包括原始数据）生命周期的所有阶段，从开始生成和记录到处理（包括转换或转移）、使用、数据保留、存档/检索和销毁（如果适用）。

3.14

测试和评估的定义方法Defined Approach to Testing and Assessment

一个定义方法包括一个固定的数据解释程序 (DIP)（例如统计、数学模型），例如统计、数学模型，应用于计算机预测、化学测试、体外试验的数据产生，即由一组确定的信息源以得出预测，而不需要专家判定。

3.15

整合测试和评估的方法Integrated Approaches to Testing and Assessment (IATA)

IATA 是一种程序性的、基于科学的化学危害表征的方法，它依赖于对现有信息的综合分析，并结合使用测试策略生成新的信息。

3.16

整合测试策略Integrated Testing Strategies (ITS)

ITS就如何评估各种类型的可用数据（包括从体外试验或测试中获得的数据）提供指导，并阐述了一些要素的其他方面，例如其他毒性数据的使用、现有和相关证据的权重分析。

3.17

检出限Limit Of Detection (LOD)

指在规定的置信限（通常为 1%）内，可以与不含该物质（空白值）区分开来的物质的最低量。

3.18

定量限Limits Of Quantification (LOQ):

包括定量下限 (LLOQ) 和定量上限 (ULOQ)，指在规定的实验条件下，以规定的可接受精度和准确度定量测定的物质的最低和最高数量。

3.19

方法终点 Method endpoint

可测量的定量特征，可作为假设的病理变化或相关生化或分子事件的指标，例如，在细胞毒性测试或体外皮肤刺激方法中测量的吸光度。

3.20

Me-too试验方法

仿制试验方法的的口语表达，它在结构和功能上与经过验证和接受的参考测试方法相似。这种测试方法将成为追加验证的候选方法。

3.21

混合物 Mixture

2种或2种以上相互不发生反应的化学物（液体或固体）的混合。

3.22

多成分物质 Multi-constituent substance

一种物质由其定量组成所定义，其中不止一种主要成分以≥10% (w/w) 和 <80% (w/w) 的浓度存在。多成分物质是制造过程的结果。混合物与多成分物质的区别在于，混合物是由两种或多种物质混合而成，没有化学反应。多成分物质是化学反应的结果。

3.23

基于实施的测试指南Performance Based Test Guideline (PBTG)

一项测试指南包含一种或多种结构上和功能上相似的体外方法。PBTG明确规定了体外方法的重要组成部分，并详细描述了一项新的体外方法应满足的特征和实施规范，以使其成为仿制的方法。

3.24

实施标准 Performance Standards (PS)

实施指南的目的是提供新的或修改的体外方法的基础，包括专利的（即版权、商标、注册）和非专利的，都可以被认为在结构和机制上与经过验证的参考方法相似，并证明对特定目的具有足够的可靠性和相关性，即符合 OECD GD 34 的原则。

3.25

阳性对照 Positive control

测试系统的单独部分，用已知的对测试系统有响应的物质给予处理。阳性对照提供了测试系统在实际的测试条件下反应的证据。阳性对照对于测试系统是终点特异性的。

3.26

预测模型 Prediction model

使用体外终点值来预测体内等效活性（即毒性程度）的方法。

3.27

熟练化学物（物质）Proficiency chemicals (substances)

标准替代方法测试指南中推荐的一组化学物，实验室在常规使用这些测试方法之前，就使用这些化学物进行测试用来证明其技术熟练程度。这些化学品可以是与OECD 测试指南相关的实施指南中包含的参考化学品的一个子集，也可以是OECD测试指南中列入的在验证研究中使用的化学品。熟练测试化学品的选择标准尽可能包括以下几点：1）代表要预测的反应范围，2）具有可用的高质量参考数据；3）涵盖方法的动态响应范围；4）在验证研究过程中被测试方法正确地预测；5）涵盖体外方法适用范围内的尽可能高的广泛性和代表性，包括物理状态、化学类别、有机官能团和结构；6）可商业获得，7）获得和处置的途径与成本不受禁止或可接受。

3.28

来源 Provenance

描述细胞系的来源和培养历史，包括它在实验室和贮存室之间的转移、操作（物理化学或遗传）、测试、检测、消除其他细胞系的污染、消除由细菌、酵母或真菌、支原体、基因型和表型特征引起的污染，并确认其身份。

3.29

相关性 Relevance

指一个程序对于特定目的是否具有意义和有用性。

3.30

可靠性 Reliability

指一个程序是否可以在实验室内和实验室之间以及随着时间的推移重复执行。

3.31

稳健性 Robustness

当试验在不同实验室进行，或在试验方法常规使用可能存在的条件宽泛的情况下，测试结果对偏离特定测试条件的不敏感性。如果测试不稳健，则很难在实验室内部和实验室之间以可重复的方式使用。

3.32

敏感性 Sensitivity

一种体外方法性能的衡量指标，描述了所有评估的测试项目在特定毒理学终点被归类为阳性的比例，这些测试项目被实际体外方法预测为阳性。

3.33

特异性 Specificity

一种体外方法性能的衡量指标，描述了所有被评估的测试项目在特定毒理学终点被归类为阴性，而实际体外方法预测为阴性的比例。

3.34

标准操作规程 Standard Operating Procedure (SOP)

一个文件化的程序，描述如何在一般计划或测试指南中不详细指定的情况下，正常地执行试验方法、测试或活动。

3.35

验证集 Validation set

一组选定的化学品，用于根据参考体外方法结果的终点值的性能评估体外方法的预测能力。验证集的测试是体外方法验证的主要部分。

3.36

溶剂对照 Vehicle/solvent control

测试系统的单独部分，其中添加了测试项目的溶剂，但没有测试项目；在体外方法的实际条件下，溶剂对照提供了所选溶剂对测试系统没有影响的证据。

3.37

实验室内评估 Within-laboratory assessment

方法验证的一个阶段，通常也称为实验室内（或内部）验证，其中来自同一实验室的一名或多名操作员独立地在不同时间执行（或运行）体外方法以确定是否体外方法符合既定标准，例如，评估实验室内再现性 (WLR)。

3.38

Z因子 Z-factor

溶剂对照和测试项目信号之间分离程度,其考虑了体外方法的动态范围以及与信号和对照测量相关的数据变化。

4 经验证和仿制的替代方法

4.1 经验证的替代方法

经验证的替代方法也称为定义性方法（DA），DA方法的组合使用旨在克服单个方法、独立方法的一些局限性，从而增加对整体结果的信心。DA的终极目标是获得相当于动物研究提供的信息，使之可用于风险识别和效力的分类。定义的方法生成的预测是基于规则的，如果认为它们适合，则可以单独使用，或者在 IATA 的背景下与其他来源的信息综合考虑。

4.2 仿制方法的开发

仿制方法开发中的内部验证目的是确定实施的等效性,优先选择但不能仅限于满足参考物质表中的物质，而应扩大测试的化学物质库。。

4.3 实施指南的三个要素

决定替代方法性能和实施的三个基本要素是测试方法的基本组件、参考化学物质的最小清单和方法准确性、可靠性的值。

4.4 实施指南的作用

4.4.1 用于追加验证

在确定一种测试方法对特定用途是有效的之后,基于该测试方法建立的实施指南可用于评估其他结构和功能相似的测试方法的有效性。所谓“追加验证”，其含义为鉴定和使用标准参考物质来评估这些类似的方法。这种快速的验证方式已成功应用于人体皮肤模型用于皮肤刺激，角膜模型用于眼刺激性的测试。无论用于评估测试方法验证状态的方式如何，所有新的或修订的测试方法都应经过正式的同行评审。

4.4.2 用于方法改进

4.4.2.1 方法改进的情况

对新开发的替代方法应定期进行审查，以确定其有效性是否发生改变。可以不定期审查和重新评估替代方法的验证状态，审查方式可以是“事件驱动的”（对应于个别需要重新评估的新的科学发现）或定期的（以确定自上次审查以来收集的新信息是否值得重新评估）。还应注意测试系统的原始属性或响应性随时间可能发生的变化，或可能影响测试方法结果的程序上的修改。对阳性、阴性和基准对照的监控有助于确定变化是否发生及其影响。

4.4.2.2 方法改进验证的实施

对已批准的测试方法的可能性改进进行评估时，应评估更改对方法性能的影响，以及更改对验证过程其他组件的影响及程度。根据变更的数量和性质、生成的数据和变更的支持文件，应该采取类似于新开发方法的验证过程，并评估可靠性和相关性。应根据具体情况确定合适的验证研究范围。替代方法的重大变化应接受新的独立同行审查，审查点包括与之前的测试比较，特定更改和改进是否提高了测试性能。此外，还应考虑对3R的影响、改进的可行性、成本效益和时间效益以及实验室能力等因素。

4.5 方法开发和实验室内验证的关系

体外方法开发和实验室内验证是连续的和相互依赖的关系。在方法开发早期阶段，通常根据体外方法的预期目的和范围选择相应的仪器和建立方法学。实验室完成方法开发和优化之后，应尽快进行实验室内验证。提供方法在实验室中运行的性能文件证据，提出确保方法在整个生命周期内进行质量监督的持续措施。并提供方法被监管或验证机构正式审查的可行性的证据。

5 可接受标准

5.1体外方法以一致性为基础

大多数体外方法通常用于定性或定量生物学反应的预测，人类或环境危害程度的预测。体外方法的运行的基础是一致性，即随着时间推移的一致和不同实验室之间的一致。

5.2 可接受标准的制定依据和内容

可接受标准应基于方法的所有关键组件和内容的历史数据而制定。标准应当给出测试系统（例如，传代次数、生长曲线、细胞复苏）和测试系统的性能（例如，阳性、阴性、溶剂对照）的定义，还应当设置分析终点的确定值（如线性、准确度、范围）和数据分析方法（如线拟合）。

5.3 可接受标准的数据要求

可接受标准主要来自历史数据中的信息。必要时，可以从验证研究或者从相关文献中补充。历史数据应使用未更改的方法，如有更改，需证明任何更改不影响数值。只有在明确、有根据的和科学合理的前提下才能拒绝数据，并且清楚准确地记录拒绝的原因。

5.4 阳性对照

5.4.1 阳性数据的产生和分布

许多情况下，阳性对照反应的可接受区间是基于实验室历史数据的统计学分析结果，数据最早产生于测试方法开发者的实验室中，随后在测试方法的验证过程积累数据，最后可能有多方实验室提供测试结果。数据通常服从正态分布，并计算得到结果的均数和标准差。结果的可接受区间一般为均数加减2个标准差。例如，牛角膜浑浊渗透性测试（BCOP）使用的是无水乙醇作为阳性对照物，它的体外评分（浊度值+15×荧光素钠渗透值）数据为51.6±6.2，那么它的可接受范围即为39.2-64.0（均数±2SD），数据来源于1171次历史测试。

5.4.2 剂量反应关系中的阳性对照

在剂量反应方法中，在准确地计算50%相对活率浓度点时，测试多个浓度使得预测值落在方法的线性动态范围内是非常重要的。阳性参照物的50%存活率（浓度）可以用于建立细胞毒性测试的可接受标准，这些测试需要稀释受试物进行不同浓度测试。这种方法对于灵敏性的上下波动有很好的识别能力。

5.5 阴性对照

建立阴性对照的可接受收标准对于确保测试系统的正常运行是很重要的，阴性对照与阳性对照同等重要，而且以相同的方式获取平均响应值（如在历史数据的 ±2 或 3 SD 范围内，或在历史数据分布的 95%置信限内）。也可以建立其他可接受标准，例如可量化数据的变异性的标准，例如：基于体外重建组织模型的方法（资料性附录A）；或者最低细胞活性水平的标准，例如皮肤致敏的树突细胞活化试验中对细胞活性的要求（OECD 指南442E）。

5.6 临界值或关键参数

可以设置可接受标准的临界值或关键参数，例如，明确规定参考或对照组的反应是否可被接受，实验结果应满足的有效数字的数量等。

6 试验设计

6.1 统计

6.1.1 统计效力

试验设计应规定每个测试条件的重复次数，包括用于参考化学物组、对照组以及测试组等的浓度水平。在体外方法开发过程中，应当使用合适的统计方法以确定实验的重复次数。例如，可以使用统计效力分析来计算理想的重复数，以测定预先设定的置信水平的不同处理之间的确切的差异。但在实际中，通常采用预设的或经验性的重复次数来替换统计效力。此外，当一个试验项目采用多个浓度时，数学模

型（例如，剂量-反应曲线）可以通过增加测试浓度和平行组的数量来拟合实验数据，但一般来说，增加浓度的数量比增加平行组的数量更好。最低平行重复数给反应参数提供了满意的可变性（例如，IC50在可接受范围内），可以作为后续验证或实施的参考。

6.1.2 非统计学因素

除了统计方面的要求外，有时要考虑实际情况的影响，如成本和重复试验的可行性。不能由于减少平行实验造成统计效力的下降。同样，需要评估独立实验的次数。例如，与使用连续细胞系的体外方法相比，使用原代细胞或组织的体外方法具有比较高的实验间变异性，后者可能需要进行更多的独立实验。

6.1.3 统计方法的作用

统计方法(例如，析因设计)在优化新的体外方法的过程中非常有用。为了获得精确、可靠和稳定的体外方法，必须评估体外方法中任何变化可能导致的结果。实验的析因设计经常用于有大量的变量需要评估的情况，因为它几乎不可能接近所有可能的组合实验。析因设计能有效地评估几个因素(自变量)的影响及可能的相互作用。预测体外方法步骤发生变化对实验参数(读数)的影响的统计方法也称为方法稳定性的评估，可用于改进方法的设计过程中，因为方法稳定性的检查可以基于多因素组合中的一个子集进行。

6.2 铺板

6.2.1 注意事项

当开发一种体外方法时，应注意尽量减少任何潜在的系统影响。基于细胞的检测多数采用细胞培养板(6、12、96或384孔板)，因此必须注意确保整个培养板(孔对孔)、培养板之间和多次运行时细胞接种、处理和测量以统一的方式进行。平板效应可能发生在外部的孔(例如，由于蒸发)，跨越列或列，甚至在实际的孔内(孔内效应)。如果测试物暴露是在对数阶段(如抑制细胞生长作为终点)，则铺板细胞密度也至关重要。在这种情况下，如果对照组在暴露时间内达到平台生长期，则测试组的效果可能被低估。另外应评估板的效应是否一致，例如，在一个完整的板上使用相同的条件和处理方式。

6.2.2 漂移

漂移可能是由于板中初始细胞接种过程中接种密度的变化，例如，细胞可能在主容器中沉淀，尤其是用于存储和接种含有细胞的悬液。此外，操作的细节，如多通道移液管吸头的松紧性，吸取容易起泡的培养基成分，可能会影响接种的准确性。无法通过技术优化解决的变异性，可能需要增加用于计算剂量反应的平行孔浓度，或需要更高的独立实验数量。

6.2.3 随机化

在细胞培养板上随机放置处理孔是一种策略，可以减少边缘效应、漂移等而产生的固有板偏倚，适合用于自动染毒装置。然而，随机化可能会引入其他不可预见的错误，如移液错误（通常是单孔移液）或数据传输和分析错误（数据可能需要重新排列以进行数据分析）。由于延长时间处理整块板反而在无意中增加了时间误差。

6.2.4 布局

根据体外方法的需要安排培养板的布局，如对照的数量、测试的浓度和所需的平行量。应当通过检查平行孔之间的可变性来控制交叉污染(例如，测试组之间)。排板设计还应考虑如何进行跨板间的比较，以便通过使用适当的参考和对照组来检查板与板之间的运行或变异性。体外实验常用的96孔板布置的示例如图1所示。

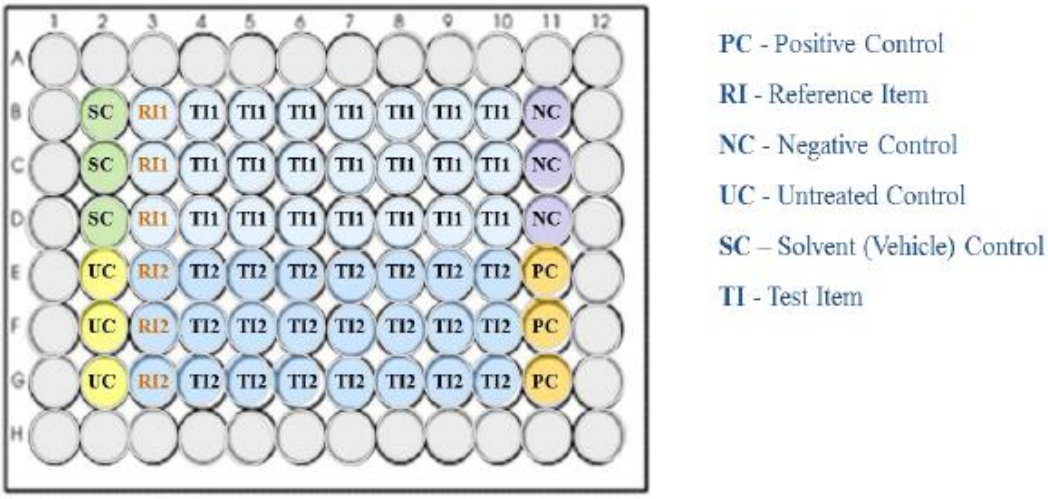

图1. 样板布局包括参考物质组，阳性和阴性对照组，溶剂组和空白对照组

6.2.5 样板

按照标准样板布局可以最大限度地减少潜在的边缘效应(由于蒸发，外孔和内孔之间的差异所致)。评价样板漂移的一种方法是规定板的左右两侧溶剂对照(SC)(图2)的差异不应超过一定的百分比，作为实验可接受的标准。例如，规定SCs的左右平均值与所有SCs的平均值的差异不超过15%。此外，某些测试或参考物质可能是挥发性的或可能通过毛细管作用污染邻近的孔，称为芯吸效应，这种情况下需要在设计板的布局时加以考虑。例如，常用的细胞裂解作用的表面活性剂Triton X可能会影响邻近孔的细胞活力，因此应在低浓度下使用，或通过在中间放置含有介质或缓冲液的孔将其与含有细胞的孔分隔。也可以在孵育前用封板膜覆盖平板，以避免挥发性测试物质的蒸发。

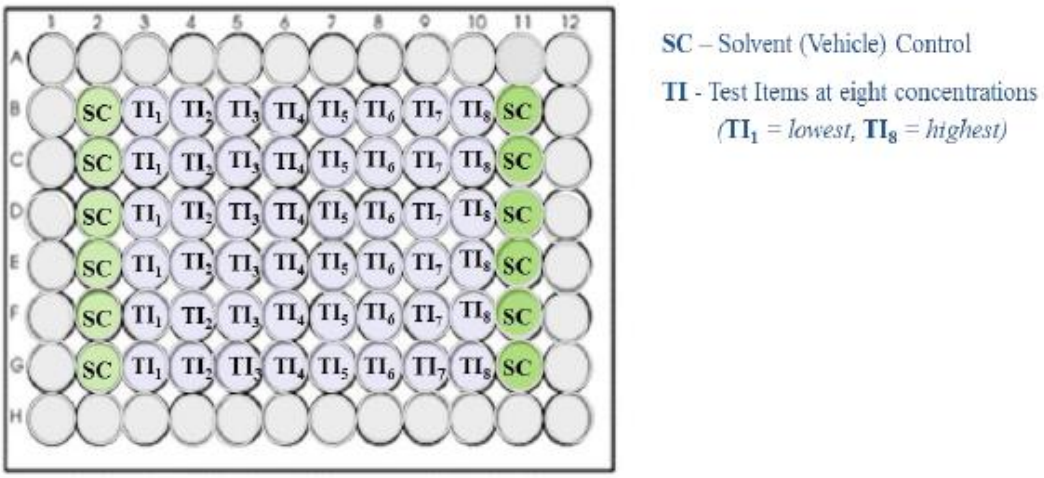

图2 系统性细胞接种错误的培养板

6.2.6 孔间效应

外侧孔与内侧孔的影响不同，不均匀的蒸发速率可能是引起变化的来源。通常最外侧的孔含有无菌的水性溶液，不同于内侧孔对照、参考或测试组，因为在打开培养箱的门时可能会发生蒸发。无细胞孔也可用于对照(例如，背景吸光度OD/荧光指数FI，检查测试物质对结果的干扰)。恒温细胞培养箱可以更好地补偿湿度的变化，减少恒温箱内的水分蒸发。应在标准操作程序中明确说明这些外侧孔的用途，如用于测试物质、参考物质或对照组。

6.2.7 积累数据

根据相关参考文献和对照组数据,在参考历史数据的基础上制定验收标准,可用于体外方法的监管。体外方法获得的数据集可以证明实验系统和分析方法的正确功能，从而证明所实施的实验的有效性。

6.3 数据分析

6.3.1 数据转换

数据转换，例如数学模型公式的规范化或拟合，应在数据获取之前定义，并应在SOP或相关研究计划中进行描述。标准化公式(检查准确性)应形成文、检查确认(如以电子格式实施时)，并公开计算中使用的对照的说明和证明。用于处理原始数据(例如，Excel表格，草稿等)的计算机脚本应该被验证和完全文档化。可以参考经合组织(OECD)关于GLP原则在计算机化系统中的应用文件。

6.3.2 数据拟合

当假定测试浓度与反应之间存在关系时，如果需要，可以拟合剂量反应曲线，以获得EC50或IC50等汇总数据。当对数据拟合数学模型(如剂量反应曲线或标准曲线)时，需要记录模型及其选择的原因。例如,当拟合剂量-反应曲线时,应将所用方程的类型(例如,四参数logistic曲线)与应用的任何约束(例如，归一化数据的最高约束为100%)、限制(例如，单调性假设)和权重(例如，逆数据不确定性)一并记录下来。此外，还应记录软件名称、版本，以及相关的参数的置信区间，例如半数有效浓度（EC50）、最高参考阳性对照浓度（RPCMax)、最高阳性对照浓度（PCMax）、半数阳性对照浓度(PC50)、10%阳性对照浓度（PC10）和相关的拟合优度参数(r平方,平方和等)用于证明模型的选择过程。在某些情况下，测试多个模型(如非单调曲线)，并在曲线拟合后选择最优拟合是可取的。

6.4 异常值检查和删除

替代方法实施时应当说明检查和删除异常值的标准，并给出理由。异常值检查方法可用Grubbs”测试(单离群值)或者使用箱线图等可视化工具(例如,马氏距离)。排除异常值时，应当考虑如何判断数据是否应当被视为异常值，是否存在未知原因所致的异常的情况。

6.5 非单调剂量和U形曲线

6.5.1 单调曲线

相对简单的剂量反应曲线是单调模型(即在整个剂量反应范围内增加或减少),但有时可以观察到非单调的剂量反应。其原因可能是由于单独产生单调剂量反应的不同效应的叠加，也有可能表明系统被扰动，例如由于化学不溶性或高浓度沉淀。

6.5.2 U形曲线

许多体外实验会形成U形曲线，例如竞争性结合数据的分析和解释的例子(图3)。因为在最高浓度下测试化学品结合百分比可能会上升而使曲线变得复杂。当测试组的浓度接近溶解度的极限时，[3H]17β-雌二醇开始产生“U形反应曲线”。出现这种U形曲线通常认为是人为影响测试条件所致，而不是测试化合物结合亲和力的实际情形。当将竞争性结合数据拟合到S型曲线时，保留这些数据点(红色圈出)会不适当地提高曲线的底部(图8.3B)，有时会导致一个测试化合物的受体(ER)结合潜力的错误分类(图3A)。如果在分析时排除一些相关数据点可以解决这一问题，依据是：与测试化合物低浓度时特定结合的平均值相比，这些数据点平均值显示出10%或更高的放射性配体结合(例如以10%的规则作为判定)。规则的制定和使用需要合理的研究报告，而且要考虑到统计学上是否合适。但应注意在数据分析时并不总是适合使用这样一个规则，特别是观察到意外和未知原因的情况时。例如有理由确定测试物质是一个非结合体，这时使用这一规则可能是适当的。

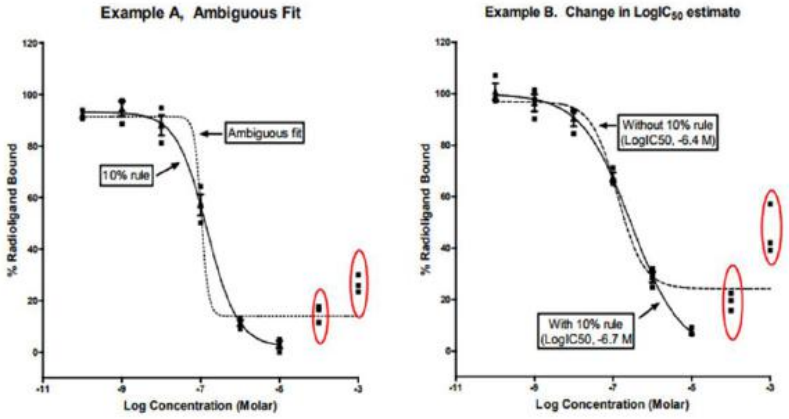

图3. 分析竞争性结合数据，采用和不采用10%规则的区别

7 测试终点的室内验证

7.1 验证的概念和要求

7.1.1 验证的要求

根据体外方法实施所处的阶段评估采取的验证方式，例如方法是否已在内部开发、是否已从其他实验室转移、是否可商用或者使用之前是否经过了全面验证。可以将方法验证分为全面验证、部分验证和交叉验证。

7.1.2 全面验证

全面验证包括方法的所有相关方面，应对所有新的内部开发方法或对现有方法进行重大更改时进行。

7.1.3 部分验证

当先前验证的方法经过微小修改（例如校准浓度范围发生变化）或已发布的方法（通常根据设施要求进行修改）被转化为实施 SOP 时，应进行部分验证。

7.1.4 交叉验证

当在同一研究或不同研究中使用两种或多种方法生成数据时要进行交叉验证。进行交叉验证时，应通过这两种分析方法分析同一组参考物质组、对照组和测试物质组。

7.1.5 验证参数

并非所有的体外方法都需要获得验证的所有方面的参数，具体取决于体外方法的目的以及定量或定性的特点。通常，定量方法至少应在适用的情况下解决方法的准确性、重现性、线性、检测限和测量范围，而对于定性方法，特异性和灵敏度是关键参数。

7.1.6 验证计划

无论采用哪种验证类型都应制定验证计划，确定验证要达到的预期目标、评估的特征参数和实施步骤。内部验证研究时，可以应用一些合格评定实验室认可的管理经验或GLP的概念，例如编制研究计划和最终报告的要求。

7.1.7 验证组织

验证应由经过培训的合格实验室人员进行，以尽量减少操作员的可变性，并且仅应使用经过校准或检定的设备，以减少设备相关问题。验证结果应与验证计划或 SOP中描述的性能特征的可接受标准进行比较，任何偏差都应记录并在总结报告（验证报告）中与验证的结果结论一起形成文件。在起草验证计划时，应注意选择适当且有意义的性能特征，不要遗漏潜在的关键特征，例如试剂多变性，因为这些将取决于被验证方法的性质和类型。

7.1.8 验证参数

定量方法的一些质量参数，如准确度参数：真实度性和精密度；不确定度；敏感性和特异性；选择性参数：干扰因素；范围和线性；检测限。定性分析方法常用的质量参数包括：敏感性、特异性；不可靠范围；假阳性和阴性率；选择性参数：干扰因素；截止限；强韧性和稳健性。应考虑不同质量参数的区别，例如，对于定量方法，灵敏度应该是一个数值，表明当分析物浓度发生变化时响应如何变化。然

而，对于通常报告结论为真假或阳性阴性的定性方法，灵敏性参数也能不适用较。这种情况同样适用于特异性、检出限、阈值和不确定性或不可靠区域等的评估。

7.2 检测限与临界值

7.2.1 数据来源

仪器在体外方法中作出的响应和给出的读数应该是已知的，并且应该是在参考物质的指定的作用浓度范围内进行评估。可以使用多种方法来确定检测限 (LOD) 和定量限 (LOQ) 。如基于视觉评价、基于信噪比、基于空白的标准偏差、基于校准曲线。LOD确定了以可接受的精度始终检测到的最低实际浓度或信号，但不一定量化。对于正态分布的数据，LOD 通常被确定为平均响应浓度 +3 SD 阴性对照范围，因为这仅提供 1% 的假阳性机会。通常根据参考物质的可接受准确度和精密度来计算LOQ。

7.2.2 信噪比

信噪比 (S/N)是指将背景噪度（观察为空白的变化）作为基线的方法，通过将来自参考物质或阳性对照组的样品的测量信号与空白样品的测量信号进行比较来计算。估计 LOD 时通常接受 S/N 比为 3，估计 LOQ 时使用 S/N 比为10。或者，可以使用信号窗口计算来确定检测的接受度。

7.2.3 定性方法

对于定性方法，检测限无法计算，因为 SD 只能在响应为数值时计算。对于这些方法，可以计算出阈值，即确定具有一定误差概率（通常为 5%）的检测所需的物质的最小浓度。 阈值通常通过在低于和高于预期临界浓度的多个水平上建立假阳性和假阴性率来确定，因此与方法的灵敏度有关。

7.3 线性和动态范围

7.3.1 概念

仪器检测器的响应可以表示为动态范围或线性动态范围。 动态范围是最大和最小浓度之比，在此范围内可以记录测量的特性（例如吸光度）。线性动态范围，即一般指检测器响应呈线性的溶质（例如参考物质）浓度范围。为了量化样品中分析物的数量，需要制作校准曲线，通常使用参考物质的稀释系列进行评估。然后根据结果绘制曲线，一般情况下是线性的，被拟合到数据上。然而，并非所有体外方法在全范围内都是线性的，因此需要在方法的范围内定义线性范围（动态范围）。对于定量测量，动态范围的边界由分析物的最低和最高浓度确定，这些浓度产生的结果由体外方法真实地产生。线性下限通常称为定量下限 (LLOQ)，线性上限称为定量上限 (ULOQ)。线性的上限可能受样品中最高可用浓度或仪器产生的信号饱和度的限制，而下限通常受仪器规格的限制。该范围通常来自线性，并通过确认该程序在应用于包含在指定分析方法范围内或极端情况下的分析物（例如，参考物质）的样品时提供可接受的线性度、准确度和精密度来确定。

7.3.2 线性拟合

如果存在线性关系，则可以采用统计方法，例如使用最小二乘法拟合回归线并计算线性回归参数（相关系数、斜率、y 截距以及残差平方和）。仅用回归计算认为一般情况下不足以建立线性和客观测试，这时可能需要拟合优度。基于拟合优度检验的相关系数 (r) 为 0.99，通常作为线性的可接受标准，但根据方法的不同，较低的 r 值也是可以接受的。 如果存在非线性关系，则可能需要在回归分析之前对数据进行数学变换。 由于线性回归更容易实现，与其他回归模型（例如，非线性）相比，一条直线的校准曲线将始终是首选。

对于某些测定/方法，如果使用者在测定/设备的范围内操作，则可以将线性以外的方程拟合为标准曲线。 但是，建议使用能够充分描述浓度-响应关系的最简单模型。选择权重和使用复杂的回归方程应该要有合理的理由。评估线性时建议至少使用 5 个浓度，但如果有其他合理理由，也可以使用其他方法。

随后，为了促进有效的体外方法转移，计算出的线性回归参数应与数据图一起提交。当超出上限时（即样品超出线性范围），可能需要对其进行稀释。 如果样品给出的结果低于线性范围的下限，则可能需要将样品制备调整为更高的浓度或更换为更灵敏的设备。

7.4 准确度和精密度

7.4.1 要求

根据体外方法是定量方法还是定性方法来对方法的准确度和精密度进行评估。定量方法的精密度是指分析物（例如参考物质）的单个测量值的接近程度，表示为变异系数 (CV)。 通常使用每个浓度至少

5 次测定和预期浓度范围内的至少 3个浓度来测量精密度，并应报告批内和批间的精密度。对于小分子，批内和批间精度不应超过 15%（LLOQ 和 ULOQ 为 20%），而对于大分子（例如肽和蛋白质），批内和批间精度不应超过 20% （LLOQ 和 ULOQ 为 25%）。总误差（即相对误差百分比和变异系数百分比的绝对值之和）不应超过 30%（LLOQ 和 ULOQ 为 40%）。

7.4.2 准确度

对于定量方法的准确度，通常使用经认证的参考物质（如果有）或通过与参考方法或其他方法的比较来确定。方法的准确度描述了通过该方法获得的平均测试结果与参考物质的实际值（或标称）值（浓度）的接近程度。通过对含有已知量分析物的验证样品进行重复分析来确定准确度。验证样品的制备应模拟研究样品的制备，并且应在几天内至少进行 6 次独立的检测。

7.4.3 精密度

精密度应报告为标称值的百分比。 在评估小分子的批内和批间准确度时，平均浓度应在每个浓度水平标称值的 15% 以内（LLOQ 和 ULOQ 为 20%）。 对于大分子，对于批内和批间精度，平均浓度应在每个浓度水平标称值的 20% 以内（LLOQ 和 ULOQ 为 25%）。

7.4.4 定性方法

定性体外方法（例如，强、弱）取决于方法的准确性和可靠性，以根据其规定的目的（例如，敏感性、特异性、阳性和阴性预测性、假阳性和假阴性率）对化学品进行正确分类。在这种情况下，可使用阈值并应考虑它们对准确性和可靠性的影响。根据与这些阈值的距离的分析，可能并不总是能得出置信界限，而更可能的情况是得出并不能确定的结果（即，既不是明确的阳性也不是阴性）。假阳性率是实际为阴性的测试物质被该方法分类为阳性的概率。

7.5 敏感性和特异性

7.5.1 定量方法

对于定量方法，灵敏度可以定义为体外方法区分测试组浓度或质量的微小差异的能力，而特异性可以定义为当存在其他物质时体外方法识别目标分析物并在适当的情况下量化的能力，即其他物质可能干扰目标分析物的鉴定/定量的程度。不可能存在证明该方法总是对特定分析物具有特异性（完全区分）。特异性取决于浓度，通常通过添加在开发该方法的样品中可能遇到的材料来确定。应在测试范围的低数值端确定特异性，并确保了解杂质、交叉反应物质等的影响。

7.5.2 定性方法

与定性方法相关的灵敏度可以定义为该方法检测真阳性率的能力，而特异性是该方法正确识别真阴性率的能力。定性方法的性能也可以用阳性预测值 (positive predictive value, PPV) 和阴性预测值 (NPV) 进行评估。PPV 是正确的阳性反应检测为阳性的比例，而 NPV 是通过体外方法检测为阴性的正确阴性反应的比例。在计算灵敏度、特异性、假阳性率、假阴性率等参数时，重要的是使用平衡的数据集（阳性和阴性化合物的数量大约相等），否则这些参数将无法反映真实情况。敏感性、特异性等水平没有被标准化的，取决于已经确认的参考物质清单。因此，这些精度参数的可接受水平的严格界限是不现实的。但是，通常情况下不接受灵敏度低于 75% 。

表3 验证中测试项目的体外方法结果的可能结果

7.6 重现性

重现性是指同一研究人员（通常是一个人）在相同操作条件下在短时间内获得的单个样品的一系列测量结果之间的接近程度，也称为批内精密度。应在例如生物统计评估之后建立表达测定重复性的最合适方法。 它通常表示为一系列测量的 % CV，但可能取决于所使用的特定体外方法和分析方法。

8 效能化学物质

8.1 说明

对于复杂的测试方法，根据测试方法的适用范围和动态范围(即响应在数据集中的分布范围)，确定一些经过验证的效能化学物质。在常规使用经验证的测试方法（如经合组织、国家和行业标准）之前，实验室可以使用这些效能化学物质来证明熟练程度。它们通常来自方法实施指南中所包含的参考化学品列表，或来自于测试方法验证研究过程中用到的化学品。

8.2 选择标准

效能化学物质的选择标准包括：1)代表要预测的反应范围，2)有高质量的可用参考数据，3)覆盖方法的动态响应范围，4)在验证研究中得到正确的预测结果，5)化学物类别代表广泛，涵盖了适用于体外方法的相关物理状态、化学类别、有机官能团和结构，6)商业上可获得，7)成本可以接受。此外，还应包括可能干扰特定体外方法指定反应的化学品，以更好地了解来自未知测试物质的潜在干扰，例如细胞毒素和细胞抑制剂、荧光化合物和荧光素酶抑制剂以及MTT干扰化学品。

8.3 使用要求

效能化学品的数量主要取决于体外方法的类型、目的和性质，并且实验结果确信是可接受的和可靠的。有些方法可能需要5种效能化学品，而有些方法则需要20种以上的化合物。体外方法使用者应在常规检测之前测试效能化学品，以达到该方法可以用于法规和监管目的。正确预测效能化学品所需的运行次数不受限制，直到满足要求为止。通过测试效能化学物质这种方式，实验室可以证明是否熟练掌握体外方法及达到的程度。也可以用于培训目的，从而证明体外方法在不同实验室间的转移和实施的能力。

9 数据密集型体外方法

9.1 验证建议

体外方法中的一部分采用了高通量和高内涵的筛选技术，这些技术可以同时对数百或数千种化合物进行处理，获得丰富的数字、描述和图像数据，对确定作用机制，促进对人体有害健康影响的预测模型的开发中应用广泛。这类体外方法可称之为数据驱动的方法。这类方法的验证与传统的体外方法无本质区别，但需要给予补充和特定说明，并在方法标准化方面形成共识，推动协调一致的验证方式。

9.2 标准化基础

图像分析方法的验证。

高通量方法的验证应当与传统的金标准比较，如果可行的话应用已建立的方法进行比较，如qPCR方法用关键的微阵列方法比较。采取标准化工作，可以实现这些体外方法的可转移性和重现性。对于许多组学技术，应促进许多关键方面的最佳实践达成共识，如样品制备和处理、实验设计和方案、数据处理、统计分析和结果解释以及质量控制等。对于一些技术，如转录组学，应关注验证研究的进展和相关机构发布指南的更新。应关注基于大数据的体外方法的应用场景及在测试策略和风险评估中的可能起到的作用，如组学技术和高通量筛查的使用场景包括：提供交叉参照的支持性证据，确定物质分类，用于开发整合测试策略等。基于“大数据”的体外方法是通过不同的技术平台和各种生化、分析和计算方法生成的，产生不同的数据类型和格式，每个关键环节和实施步骤应做到一定程度的标准化。以确保结果的重现性、可靠性、可接受性和正确性。

资料性附录 A 皮肤模型刺激性测试的实施指南

资料性附录 B 基于转染报告基因的角质细胞氧化反应活化的皮肤致敏方法的实施指南

附　录　A 皮肤模型刺激性测试的实施指南

（资料性）

A1前言

以OECD测试指南439最初验证认可的方法作为基础的改进方法，在大规模应用于测试化学物之前，应先使用参考化学进行评估以确定其可靠性和预测能力。这些化学物质来自于皮肤刺激体内动物实验的参考刺激物和非刺激物。以确保改进的方法能够根据联合国GHS制度正确地区分刺激性和非刺激性化学品。仿制的或改进的测试方法与验证的参考方法（VRM）相比，具有相同或更好的再现性、敏感性、特异性和准确性。

A2 测试方法的基本组件

科学有效的测试方法（valid test methods,VRM）由基本结构、功能和程序三个要素组成，对于已验证的方法和结构功能上与之相似或改进的方法都是一致的。通过体外方法的三个要素，可以明确体外方法的独特性质、关键程序细节和质量控制措施。

A3 可改变的值或实验步骤

具体到每个体外方法，可能存在特定参数（例如附表A1和附表A2）或实验步骤有所不同的情况，拟提议的仿制或修改的方法应当提供相应的值或步骤的修改。

A4 重建表皮的特征要求

重建表皮应使用非转化的人角质形成细胞，RhE模型是在多孔的合成膜的插入皿中制备的，营养物质可以通过该膜进入细胞。在功能性角质层下应存在多层有活力的上皮细胞（基底层，棘皮层，颗粒层）。将测试化学品局部应用于三维RhE模型，模型表面应与空气直接接触，以模拟与体内情况相似的暴露。角质层应多层包含必需的脂质特征，以产生功能强大的功能性屏障，并具有抵抗细胞毒性化学物质（例如抗氧化剂）快速渗透的稳健性。常用的测试剂为十二烷基硫酸钠（SDS）或TritonX-100。屏障功能应得到证明，并通过确定的参数值进行判定。例如检测化学物质在固定暴露时间后将组织活力降低50％（IC50）的浓度，或通过在指定的固定浓度下确定细胞生存能力降低50％（ET50）所需的暴露时间来评估。RhE模型的密闭特性应防止测试化学物质绕过角质层到达存活的组织，这将导致错误的皮肤暴露。RhE模型应不受细菌，病毒，支原体和真菌的污染。

A5 组织生存力

组织活性的定量测定采用MTT方法。RhE的活细胞组织构建体可以将活体染料MTT还原为蓝色MTT甲瓒沉淀，然后使用异丙醇（或类似溶剂）从组织中提取的光密度（OD）单独的萃取溶剂应足够小，即OD<0.1。可以使用标准吸光度（OD）测量或HPLC/UPLC-分光光度法对提取的MTT甲瓒进行定量。RhE模型使用者应确保使用的每批RhE模型均符合阴性对照的定义标准。RhE模型供应商应确定阴性对照OD值的可接受范围（上限和下限）。附表A1给出了RhEVRM阴性对照OD值的可接受范围。HPLC/UPLC的实验人员应使用附表A1中提供的阴性对照OD范围作为阴性对照的可接受标准。应当证明，用阴性对照处理的组织在测试暴露时间段内在培养中是稳定的（提供相似的活性测定）。

附表A1：VRM的阴性对照OD值的可接受范围

	接受下限	接受上限
EpiSkinTM (SM)	≥ 0.6	≤ 1.5
EpiDerm™ SIT (EPI-200)	≥ 0.8	≤ 2.8

A6 屏障功能

角质层及其脂质成分应足以抵抗快速渗透，可以通过IC50或ET50估算某些细胞毒性的基准化学品（例如SDS或TritonX-100）进行评估（附表A2）。

A7.形态学

应该对RhE模型进行组织学检查，以证明其具有人类多层状的人表皮样结构，包含基础层，棘状层，颗粒层和角质层，并显示出类似于人表皮脂质分布的脂质分布。

A8 重现性

测试方法的阳性和阴性对照的结果应证明 随着时间的推移，测试方法的可重复性。

A10 质量控制

使用的RhE模型符合确定的生产标准，其中与生存力、屏障功能、和形态最相关。RhE模型供应商应确定IC50或ET50的可接受范围（上限和下限）。附表A2给出了VRM的可接受范围 。对于任何新的类似或改进的测试方法，应提供足够的范围。这些可能会因具体的测试方法而异。RhE模型供应商应提供证明符合所有生产发布标准的数据。只有用满足所有这些生产质量标准的组织所产生的结果才能被接受，以便可靠地预测刺激分类。

附表A2：VRM的QC批次标准

	接受下限	接受上限
EpiSkinTM (SM)(SDS 处理 18 h)	IC50 = 1.0 mg/ml	IC50 = 3.0 mg/ml
EpiDerm™ SIT (EPI-200) (1% Triton X-100)	ET50 = 4.0 hr	ET50 = 8.7 hr

A11 平行测试

每种测试化学品和对照品中至少应使用三个组织重复样本。每次运行。对于液体和固体化学品，应使用足够量的测试化学品以均匀覆盖表皮表面，同时避免不受限制加量（例如，对于VRM，最低剂量为26μL/cm2或mg/cm2）。尽可能将固体作为细粉进行测试。针对不同的RhE模型优化了暴露时间和温度，并与RhE模型各自不同的属性（例如屏障功能）相关。此外，组织活性测量不应在暴露于测试化学品后立即进行，而应在去除测试物质后经过组织足够长的孵育期后进行。这个时期的组织既可以从弱的细胞毒性作用中恢复，也可以出现明显的细胞毒性作用。研究发现处理后42小时的孵化期最适合VRM。在每次运行中应同时使用阴性对照（NC）和阳性对照（PC）。证明组织的活性（使用NC）和敏感性（使用PC）在定义的历史接受范围内。并发阴性对照还提供基线（100％组织活力）以计算用测试化学品处理过的组织的相对活力百分比。建议用于VRM的PC为5％SDS水溶液。建议的VRMNCs为磷酸盐缓冲盐水（PBS）。

A12 MTT干扰

A12.1 MTT反应原理

MTT是一种定量测定组织生存力的方法，通过比色法测试活体染料MTT还原为蓝色的甲瓒产物的量，反应细胞的活性。因此，与MTT染料还原反应类似，或者与甲瓒光吸收类似，或者自身有颜色的物质的测试可能会受到干扰。

A12.2测试MTT还原剂

将每种测试化学品添加到新制备的四甲基偶氮唑盐（MTT）溶液中。如果混合含有试验化学品的MTT溶液变蓝或变紫，则认为试验化学品直接还原了MTT，应在无活性的重组人皮肤模型组织上进行进一步的检查。这种额外的检查采用无活性的组织，这些组织仅具有残余的代谢活性。将可能还原MTT的化学物质应用于至少两个无活性的平行组织完成整个实验，得到非特异性的MTT还原(NSMTT)。无论进行多少次平行测试，对于每一种测试化学品只需一个NSMTT对照。然后计算真实的细胞活性，即活组织暴露于MTT还原剂时获得的组织活性百分比减去死亡组织暴露于相同MTT还原剂时获得的非特异性 MTT 减少百分比。相对于校正试验同时运行的阴性对照计算(%NSMTT)。独立于使用标准吸光度测量或高效液相色谱/UPLC分光光度法程序。

A12.3 测试化学品的光学性质

测试化学品的光学性质可能对MTT产生干扰，导致对组织活力的错误估计。通过将MTT直接还原成蓝色甲瓒，或者如果受试化学物自然吸收或由于处理过程导致吸收甲瓒，原因是与甲臜相同的OD范围（即570±30nm，主要是蓝色和紫色化学品）。这种情况下，测试前应进行预检查，以识别潜在的直接MTT还原剂或颜色干扰化学物质。建立相应的标准化程序，并作为SOP的一部分。例如非特定的MTT降低（NSMTT）控制和非特定的颜色（NSC）控制。当测试物质不能通过冲洗从组织中完全清除时，有可能穿透表皮，造成MTT生存力测试时受试物存在于组织中，干扰作用尤为突出。对于有色的化学物或与水或异丙醇接触而变成有色的测试化学药品，由于对MTT分析的干扰太强（例如，在570±30nm处有强吸收）而与标准吸光度（OD）测量不兼容，可以采用HPLC/UPLC分光光度法测量MTT甲瓒。

A12.4 测试过程

为了确定直接的MTT还原剂，应将每种测试化学品添加到新鲜制备的MTT中溶解。如果含有测试化学物质的MTT混合物变成蓝色/紫色，则认为该测试化学物质会直接降低MTT，因此应独立于标准吸光度（OD）测量或HPLC对无活性的RhE组织进行进一步的HPLC/UPLC分光光度法程序。使用无活性的组织（例如通过暴露于低温下“冷冻杀死”的组织或通过其他方式），该组织仅具有残留的代谢活性，但可以与存活组织相似的方式吸收和保留测试化学品。将可能还原MTT的测试化学品施加到至少两个平行进行完整测试程序的无活性的组织。真实组织活力的计算方法是：暴露于MTT还原剂的活组织获得的组织活力百分比减去暴露于同一MTT还原剂的无活性组织获得的非特异性MTT降低百分比，相对于同时进行的阴性对照计算得出的纠正MTT（％NSMTT）。

A12.5 识别有色化学品的测试

测试化学物与水或异丙醇接触时是否产生颜色，并决定是否需要额外的对照，应对水(暴露期间的环境)和/或异丙醇(提取液)中的测试化学品进行光谱分析。如果水和/或异丙醇中的试验化学品吸收570±30纳米范围内的光，则应进行进一步的着色剂对照，或者，应使用高效液相色谱/UPLC分光光度法程序，在这种情况下，不需要这些对照。进行标准吸光度(OD)测量时，将每种干扰性有色测试化学品应用于至少两个活组织复本，这些复本经过整个测试程序，但在MTT孵育步骤中用培养基而不是MTT溶液孵育，以产生非特异性颜色(NSCliving)对照。由于活组织固有的生物可变性，需要在测试有色测试化学品的同时进行对照组测试，如果进行多次测试，则每次测试(每批实验)都需要进行独立的对照测试。然后，将真实的细胞活性计算为活组织暴露于干扰试验化学品并与 MTT 溶液一起孵育所获得的组织存活率百分比减去活组织暴露于干扰。

A12.6 检测方法的干扰

高效液相色谱/UPLC分光光度法也可用于所有类型的测试化学品(有色、非有色、MTT还原剂和非MTT还原剂)，以测量MTT甲臜。前述NSMTT对照不仅可以校正被测化学物质潜在的直接MTT减少，而且可以校正被测化学物质残留在皮肤内导致的的颜色干扰。由于NSCliving对照已经校正了由于活体组织保留测试化学物质而产生的颜色干扰，因此这可能导致对颜色干扰的双重校正。为了避免这种情况，需要对无活性组织中的非特定颜色（NSCkilled）进行第三次对照。具体操作是将测试化学品施加到至少两个无活性组织上，进行完整测试，但在MTT孵育步骤中与培养基而不是与MTT溶液一起孵育。每个测试化学品只需一个NSC的对照就足够了。无论执行的独立测试/运行的次数如何，都应与NSMTT对照同时进行，并在可能的情况下使用相同的组织批次进行。真实组织活力的计算方法是：暴露于测试化学物质的活组织获得的组织活力百分比减去％NSMTT减去％NSCliving，再加上暴露于干扰试验化学物质并与不含MTT的培养基一起孵育的杀死的组织获得的非特异性颜色百分比，相对于阴性对照进行的校正计算值与校正的测试同时进行（％NSCkilled）。

A12.7 注意事项

用HPLC/UPLC分光光度法时，不需要使用NSCliving或NSCkilled的对照，不受测试化学品的影响。但是，如果怀疑该测试化学品会直接降低MTT或其颜色（本身或与水混合时）的颜色（直接或间接与水混合使用）妨碍评估直接降低MTT的能力，则应使用NSMTT对照。当使用HPLC/UPLC分光光度法测量MTT甲瓒时，将组织活力百分比计算为相对于同时阴性对照获得的MTT甲瓒峰值，将活体组织暴露于测试化学物质而获得的MTT甲瓒峰面积的百分比。对于能够直接降低MTT的测试化学药品，将真实的组织活力计算为将活组织暴露于测试化学药品所获得的组织活力百分比减去％NSMTT。最后，应该注意的是，在极少数情况下，直接MTT减少剂或MTT减少剂也会产生颜色干扰，并且在处理后保留在组织中，如果导致OD读数错误，VRM可能无法对其进行评估（使用标准OD测量值）或被测组织提取物的峰面积（使用UPLC/HPLC-分光光度法）超出了分光光度计的线性范围。

A13 对照

对于每次运行，用阴性对照处理的组织应表现出OD值，反映组织的质量。阴性对照反应组织运输、接收和所有试验步骤完成之后的总的误差，不应超出历史确定的范围。同样，用PC处理过的组织的活性应在历史确定的平均组织活性的范围内（相对于阴性对照），从而反映出组织在测试方法条件下对刺激性化学物质作出反应的能力。测试化学品和对照物质的平行组织之间的变异性应落在同样的根据历史值确定的可接受范围内（例如，VRM的SD≥18）。如果实施中包含的NC或PC不在范围内，如果测试化学品

的组织平行之间的变异性超出了可接受范围，则应重新测试。重要的是，不合格运行的频率增加可能表明测试系统（例如RhE本身质量）或处理（例如运输，SOP执行）方面的问题。因此，应仔细监测验证研究中不合格运行的发生，并且应报告所有不合格运行。

A14 结果解释和预测模型

A14.1 结果统计

每种测试化学品获得的OD值应用于计算以下物质的百分比：阴性对照的组织活性设为100％。在使用HPLC/UPLC分光光度法的情况下，将组织活性百分比计算为将活组织暴露于测试化学药品中获得的MTT甲瓒峰面积相对于同时阴性对照获得的MTT甲瓒峰面积的百分比。

A14.2预测模型

根据OECD439的预测模型，皮肤刺激性预测的VRM的判定值：如果暴露和孵育后的皮肤活性≤50％。根据UNGHS将测试化学品标识为需要分类和标记。由于VRM不能区分GHS类别1和类别2，因此需要进一步的皮肤腐蚀信息来决定测试化学品的最终分类。如果发现测试化学品无腐蚀性（例如，基于TG430、431或435），并且在暴露和处理后孵育后组织活性≤50％，则测试化学品根据GHS分类为类别2可以认为对皮肤有刺激性。如果暴露和孵育后VRMS中的组织活性>50％，则可以将测试化学品视为不需要分类和标记（UNGHS无分类）。

A15 最小参考化学物质清单

A15.1 参考化学品的代表性

参考化学物质可用于确定体外方法的可靠性和预测能力。一种提议的仿制或修改的测试方法，如果证明其在结构和功能上与VRM足够相似，或者只是对VRM的微小修改，该方法至少应与VRM方法相同或更好。附表A3中列出20种参考化学品代表不同化学类别（即基于官能团的化学类别），并且代表了Draize皮肤刺激性评分的所有范围（从非刺激性到强烈刺激性）。清单中包括10种UNGHS2类化学品和10种非分类化学品，其中3种是可选的UNGHS3类化学品。类别3可以认为是“无类别”。

A15.2 使用说明

参考物质表中的20种化学品可以满足仿制或优化方法的可靠性和预测能力的确认。应尽可能避免使用清单外的参考化学品。除非出于其他合理原因无法使用，这种情况下，选择参考化学品应遵循8.2的标准，并提供足够的体内参考数据，例如优先来自预验证后优化过程中使用的测试化学品，或来自VRM的验证研究。

附表A3：参考化学品的最低清单

化学品	CAS 号	物理状态	体内评分	基于体外的VRM*分类	基于体内结果的 UN GHS 分类
非分类化学品					
1-Bromo-4-chorobutane	6940-78-9	Liquid	0	Cat.2	No Cat.
Diethyl phthalate	84-66-2	Liquid	0	No Cat.	No Cat.
Naphthalene acetic acid	86-87-3	Solid	0	No Cat.	No Cat.
Allyl phenoxy-acetate	7493-74-5	Liquid	0.3	No Cat.	No Cat.
Isoprpanol	67-63-0	Liquid	0.3	No Cat.	No Cat.
4-Methyl thio benzaldehy de	3446-89-7	Liquid	1	Cat.2	No Cat.
Methyl stearate	112-61-8	Solid	1	No Cat.	No Cat.
Heptyl butyrate	5870-93-9	Liquid	1.7	No Cat.	No Cat. (Optional Cat.30)
Hexyl salicylate	6259-76-3	Liquid	2	No Cat.	No Cat.

					(Optional Cat.30)
Cinnamaldehyde	104-55-2	Liquid	2	Cat.2	No Cat. (Optional Cat.30)
分类化学品					
1-Decanol2	112-30-1	Liquid	2.3	Cat.2	Cat.2
Cyclamen aldehyde	103-95-7	Liquid	2.3	Cat.2	Cat.2
1-Bronmohexane	111-25-1	Liquid	2.7	Cat.2	Cat.2
2-Chloromethyl-3,5-dimethyl-4-methoxypyridine HCL	86604-75-3	Solid	2.7	Cat.2	Cat.2
Di-n-propyl disulphide2	629-19-6	Liquid	3	No Cat.	Cat.2
Potassium hydroxide(5% aq.)	1310-58-3	Liquid	3	Cat.2	Cat.2
Benzenethiol,5-(1,1-dimethylethyl)-2-methyl	7340-90-1	Liquid	3.3	Cat.2	Cat.2
1-Methyl-3-phenyl-1-piperazine	5271-27-2	Solid	3.3	Cat.2	Cat.2
Heptanal	111-71-7	Liquid	3.4	Cat.2	Cat.2
Tetrachloroethylene	127-18-4	Liquid	4	Cat.2	Cat.2

A16 可靠性和预测能力的确认

A16.1 验证要求

为了建立仿制或优化方法的可靠性和预测能力（即敏感性，特异性和准确性），以使能够在多个实验室使用，附表A3中列出的所有20种参考化学品都应至少在三个实验室中进行测试。但是，如果建议的测试方法仅在某一个实验室中使用，则不需要多个实验室测试来进行验证。在每个实验室中，所有20种参考化学品均应在足够间隔的时间点用不同批次的组织进行三个独立的运行测试。每次运行应至少由三个同时平行测试的组织组成，分别用于直接测试MTT降低和/或颜色干扰的每种测试化学品，阴性对照，阳性对照和适应性对照。

A16.2 验证的统计

实验室内可重复性，实验室间可重复性的计算，应获得测试方法的准确性、敏感性和特异性值，如下所述： 只有来自完整运行实验室的数据才有资格在实验室内部以及实验室之间的变异性和预测能力（准确性）中进行计算。在每个参与实验室中，每种参比化学品的最终分类应使用完整运行实验室中不同运行期间的组织活性平均值得出。只有在所有参与实验室获得的具有完整运行顺序的化学品的数据才有资格计算实验室间变异性的测试方法。预测能力（敏感性，特异性和准确性值）的计算应基于不同参与实验室针对20种参考化学品获得的单个实验室预测进行。在这种情况下，运行顺序由一个实验室针对一种测试化学品进行的三个独立运行组成。完整的运行顺序是来自一个实验室的一种测试化学品的运行顺序，其中所有三个运行均有效。这意味着任何单个无效运行都会使含有三个运行的整个运行无效。

A17 实验室内可重复性

实验室内可重复性的评估应在一个实验室中进行。目的是在20种参考化学品的不同独立测试运行中，获得预测（UNGHS第2类和无类别）的一致性（≥90％）。

A18实验室之间的可重复性

如果建议的测试方法对实验室之间的可重复性进行评估不是必不可少的仅在单个实验室中使用。对于要在实验室之间转移的方法，至少三个实验室之间在20种参考化学品的不同，独立测试运行中获得的预测的一致性（UNGHS第2类和无类别）应等于或大于（≥）80％。

A19 预测能力

仿制或经修改的体外方法的预测能力（敏感性，特异性和准确性）应等于或优于VRM得出的目标值。灵敏度应≥80％。特异性应≥70％。准确性应≥75％。需要注意的是，1-癸醇（尽管体内兔试验中被鉴定为刺激物，但已知对人体无刺激性，1-decanol是临界参考化学品，而二正丙基二硫化物是VRM假阴性化学品。由于RhE模型基于人类细胞，因此它们可以将这些参考化学品预测为无刺激性（UNGHS类别）。

附表A4

Sensitivity	Specificity	Accuracy
≥ 80%	≥ 70%	≥ 75%

A20 验证可接受标准

验证不可接受的情况包括：不符合测试接受标准（不合格的测试），或由于其他原因（例如技术原因），或由于同时进行的阳性、阴性对照失败。为了补充丢失的数据，最多允许再进行两次运行（“重新测试”）。更准确地说，在重新测试的情况下，还必须同时测试阳性和阴性对照物质，因此每个实验室中的每种参考化学品最多可以进行两次额外的运行。不合格的测试应形成文件并报告。重要的是，每个实验室每个参考化学品的合格测试不得超过三个。更多的测试并不必要的。所有测试使用过的皮肤组织均应记录。应记录不可接受的测试/运行的程度，并应分析可能的原因。即使在重新测试之后，每个参与实验室中的每种参考化学物质也没有获得最少三个有效运行次数的数据，结果是数据矩阵不完整。在这种情况下，为了满足基于PS的验证研究的目的，可以考虑接受以下三个标准： 所有20种参考化学品应在一个实验室中至少具有一个完整的运行顺序；每个至少三个参与实验室，应至少具有完整运行顺序的85％（对于20种参考化学品：每个实验室允许3个无效运行顺序）；.至少需要来自三个实验室的所有运行序列中至少有90％是完整的（在三个实验室中测试了20种参考化学品：总共允许6个无效运行序列）。

附　录　B 基于转染报告基因的角质细胞氧化反应活化的皮肤致敏方法的实施指南

（资料性）

B1前言

以OECD测试指南最初验证认可的方法作为基础的改进方法，在大规模应用于测试化学物之前，应先使用参考化学进行评估以确定其可靠性和预测能力。这些化学物质来自于眼刺激体内动物实验的参考刺激物和非刺激物。以确保改进的方法能够根据联合国GHS制度正确地区分刺激性和非刺激性化学品。仿制的或改进的测试方法与验证的参考方法（VRM）相比，具有相同或更好的再现性、敏感性、特异性和准确性。

B2 测试方法的基本组件

科学有效的测试方法（valid test methods,VRM）由基本结构、功能和程序三个要素组成，对于已验证的方法和结构功能上与之相似或改进的方法都是一致的。通过体外方法的三个要素，可以明确体外方法的独特性质、关键程序细节和质量控制措施。

B 3 标准化替代方法可改变的值或实验步骤

B3.1 可改变的情况

具体到每个体外方法，可能存在特定参数（例如附表A1和附表A2）或实验步骤有所不同的情况，拟提议的仿制或修改的方法应当提供相应的值或步骤的修改。

B2.2 细胞

应使用与皮肤致敏过程相关的转基因人类细胞系（例如，角质形成细胞，树突状细胞和其他相关细胞），并具有稳定的荧光素酶报告基因插入功能。

B3.3 报告基因

报告基因应处于人AKR1C2基因的ARE元件或其他ARE元件的控制之下。

B3.4 ARE元件

如果使用包含替代ARE元素的细胞系，则应证明所选ARE元素对Nrf2的特定依赖性以及报告基因活性的Nrf2依赖性（例如，通过使报告细胞系中的Nrf2活性暂时或永久失活并通过验证由于Nrf2表达减少或丢失，致敏剂诱导的荧光素酶活性降低）。

B3.5 萤光素酶表达

应该证明该细胞系具有萤光素酶报道基因的稳定插入。应根据性能选择克隆，例如基于荧光素酶诱导的光输出的最佳信噪比，以及当用非致敏剂处理细胞时荧光素酶诱导的最大动态范围。

B3.6 细胞培养条件

应该确定最佳的细胞接种数量和培养基成分（例如DMSO，FCS），以确保皮肤致敏剂（包括弱致敏剂）获得显著的萤光素酶诱导作用。

B3.7 萤光素酶检测

荣光素酶报告基因活性的测量和所用的萤光素酶底物应具有足够的光输出，以确保足够的灵敏度和低变异性。

B3.8 应评估细胞的细胞毒性。

B4参考物质清单

B4.1 参考物质

参考物质用于确定仿制或修改的试验方法的可靠性和相关性，经验证在结构和功能上与VRM足够相似，或对VRM有微小改进的测试方法，是否与VRM相当或更好。表1列出的推荐参考物质包含了体内皮肤致敏作用的全部范围的代表物质，它们通过各种机制起作用，并根据其官能团代表了不同的化学类别。此列表中包括的物质包括皮肤非致敏剂和皮肤致敏剂，包括LLNA EC3值确定的各种作用类别（弱，中，强和极端）。这些物质选自VRM验证研究中使用的物质，并在欧洲验证机构EURL ECVAM进行的独立同行评审中进行了评估。

B4.2 最小参考物质清单

附表B1中列出的20种参考物质代表了应用于区分皮肤致敏剂和非致敏剂的可靠性和相关性的最小物质数量。用于评估仿制或改进试验方法的预测能力（PC）和实验室间重现性（BLR）时，应区分皮肤

致敏剂和非致敏剂（12种具有不同效力的皮肤致敏剂和8种非致敏剂）。用于评估实验室内重现性（WLR）时，应根据附表B1所列的20种参考物质中的12种（斜体，包括8种具有不同效力的皮肤致敏剂和4种非致敏性剂）进行评估。在无法使用所列物质的情况下，可以使用其他具有足够体内参考数据的物质，主要是从VRM验证研究中使用的物质中获取。如果需要，可以将代表其他化学类别并具有足够体内参考数据的其他物质添加到参考物质列表中，以进一步评估建议的测试方法的相关性。

附表B1：荧光素酶试验方法参考物质列表

说明：用于评估实验室内再现性（n=12，WLR）；用于评估实验室间再现性（n=20，BLR）；用于评估预测能力（n=20）

参考物质a	CASRN	物理性质	体内分类b	体外预测
基于体内结果的非致敏剂				
4-甲氧基苯乙酮c	100-06-1	固体	非致敏性	阳性
甘油	56-81-5	液体	非致敏性	阴性
异丙醇	67-63-0	液体	非致敏性	阴性
水杨酸	69-72-7	固体	非致敏性	阴性
氯苯	108-90-7	液体	非致敏性	阴性
乳酸	50-21-5	液体	非致敏性	阴性
水杨酸甲酯	119-36-8	液体	非致敏性	阴性
磺胺	63-74-1	固体	非致敏性	阴性
基于体内结果的皮肤致敏剂				
二甲基丙烯酸乙二醇酯	97-90-5	液体	皮肤致敏剂（弱刺激性）	阳性
苯甲酸苯酯	93-99-2	固体	皮肤致敏剂（弱刺激性）	阴性
柠檬醛	5392-40-5	液体	皮肤致敏剂（中等刺激性）	阳性
4-丙烯基-2-甲氧基苯酚	97-54-1	液体	皮肤致敏剂（中等刺激性）	阳性
甲基二溴醇戊二腈	35691-65-7	固体	皮肤致敏剂（强刺激性）	阳性
对苯二胺	106-50-3	固体	皮肤致敏剂（强/极端刺激性）	阳性
2，4-二硝基氯苯	97-00-7	固体	皮肤致敏剂（极端刺激性）	阳性
4-硝基苯基溴化物	100-11-8	固体	皮肤致敏剂（极端刺激性）	阳性
丁香油酚	97-53-0	液体	皮肤致敏剂（弱刺激性）	阴性d
二甲苯并噻唑	149-30-4	固体	皮肤致敏剂（中等刺激性）	阳性
硫酸甲基苯酚	55-55-0	固体	皮肤致敏剂（强刺激性）	阳性
氧氮杂茂酮	15646-46-5	固体	皮肤致敏剂（极端刺激性）	阳性

a.根据测试目的和皮肤致敏程度排序，

b.体内类别基于LLNA测试方法的EC3值（10<弱；1<中等<10；0.1<强<1；极端<0.1）。

c.此参考物质无实验室之间变异性KeratinoSensTM数据

d.对此参考物质结果实验室间未100%一致。

B4.3 参考物质选择标准

参考化学品选择标准包括：（I）代表皮肤致敏效力范围的物质（例如,在VRM中测试的非致敏剂、弱、中、强和极端皮肤致敏剂）；（II）尽可能反映BLR和PC的VRM的性能特性；（III）具有明确的化学结构；（IV）包括各种作用机制（包括亲电剂、氧化剂、加合物形成、迈克尔受体、希夫碱形成、酰基转移化学品、芳基电腈、电腈-h-极性化学）；（V）包括基于其有机官能团的各种化学类别；（VI）诱导体内参考试验方法的最终结果；（VII）可商品供应；（VIII）易于处理。

B5 定义的可靠性和准确性值

B5.1 验证要求

为了确定仿制或修改的体外皮肤致敏试验方法的可靠性和相关性：应测试附表B1中列出的所有参考物质。应在至少三个实验室中的每个实验室对20种参考物质进行测试。为了评估实验室内的可重复性，每个参与实验室应测试表1中以斜体标识的12种参考物质，以产生三轮合格的实验，即每个实验室中得出三个预测。其余8种参考物质应由每个实验室在一个单独的合格实验中进行测试，即每个实验室得出一个预测，三个实验室的总计三个预测。最后，应使用所有20种参考物质评估预测能力。如果获得一致的结果，则一个合格的实验至少需包含两个合格的独立重复实验。如果前两个重复的结果不一致，则至少需包含三个合格的独立重复实验。对于每个测试浓度，每个测试均由三个平行重复组成。阴性（溶剂）对照和阳性对照应在每次重复中与测试化学品同时进行三次重复测试。

B5.2 实验室内重现性、实验室间重现性、准确性、灵敏度和特异性值的计算：

B5.2.1 实验室内可重复性（WLR）

应基于每个参与实验室对附表B1中用斜体字标出的12种参考物质的子集所进行的分类的一致性，使用三个合格的实验来计算。

B5.2.2 实验室间可重复性（BLR）

应基于至少三个参与实验室对表1所列的20种参考物质获得的分类一致性进行计算。对于每个实验室将为其生成三种分类（用于WLR评估）的12种物质，应根据获得的三个预测的模式为每个实验室得出一个单一的最终分类。然后将这些单一的最终分类用于BLR评估。

B5.2.3 准确度

准确度值的计算应使用至少三个实验室使用20种参考物质进行的所有合格实验进行。计算应基于每个合格实验获得的预测。

B5.2.4 数据质量要求

应使用加权计算来考虑以下事实：20种参考物质将进行不同数量的实验，即用于BLR和WLR评估的12种物质将分别进行9种实验，而用于BLR评估的8种物质每个只会进行三个实验。总而言之，在计算中应将每种物质（从所有参与实验室获得）的每个合格实验的最终结果作为独立分类，并应用校正因子，以便所有物质在计算中的权重相等。因此，每种物质的肯定和否定预测应除以该物质的预测总数，以使每种物质在计算中的最终权重为1。其中，合格实验包括一个实验，该实验包含必要数量的重复，这些重复符合SOP中定义的阴性和阳性对照的接受标准，以及OECD 442D“皮肤致敏性测试指南：ARE-Nrf2荧光素酶测试方法”的要求。否则，该实验被视为不合格。

B6 实验室内可重复性

对实验室内可重复性的评估应显示预测的一致性。在每个参与实验室的12种推荐参考物质（表1中以斜体显示）的三个不同的独立合格实验中获得的（阳性与阴性）数据值应等于或高于（≥）80.0%（基于验证数据，KeratinoSensTM的实际值为85.7%）。

B7 实验室间结果可重复性

对于类似或改进的试验方法，20种推荐参考物质（附录B表1所示）的预测一致性（阳性与阴性）应等于或高于80.0%（基于附录B表1中所示的19种参考物质KeratinoSensTM的实际值为94.7%，以及基于验证数据集的KeratinoSensTM的实际值85.7%，排除无法获得BLR上KeratinoSensTM数据的4-甲氧基苯乙酮）。

B8 预测性

仿制或改进的体外方法的准确性、灵敏度和特异性应与VRM相似或更好。使用附录B表1中列出的20种参考物质获得的准确性、敏感性和特异性均应≥80.0%。对于KeratinoSensTM，基于20种参考物质并

使用加权计算的实际结果：87.0%的准确性，86.7%的敏感性和87.5%的特异性。不应将任何强或极端的致敏剂低估为非致敏剂，除非可以给出明确的理由。

参 考 文 献

化妆品安全技术规范

OECD. 2005. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, OECD Series on Testing and Assessment No. 34, OECD, Paris.

OECD.2018. Guidance Document on Good In Vitro Method Practices (GIVIMP), OECD Series on Testing and Assessment, No. 286, OECD Publishing, Paris.

OECD.2018.Guiding Principles and Key Elements for Establishing a Weight of Evidence for Chemical Assessment Series on Testing and Assessment No. 311, OECD Publishing, Paris.

OECD 492 重建人角膜样上皮(RhCE)测试方法（Test method for reconstructed human corneal epithelium (RHCE)）

OECD 442D 体外皮肤敏感试验：针对皮肤敏化不良结果通路上角质细胞活化的关键事件检测（In Vitro Skin Sensitisation Assays addressing the AOP key event on keratinocyte activation）

OECD 430 体外皮肤腐蚀：经皮电阻测试方法（TER））（In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test Method (TER)）

OECD 431 体外皮肤腐蚀:重建人表皮(RhE)试验方法（In Vitro Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method ）

OECD 439 In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method（体外皮肤刺激:重建人体表皮试验方法）

OECD 492 重建人角膜样上皮(RhCE)测试方法（Test method for reconstructed human corneal epithelium (RHCE)）

OECD. Series on Testing and Assessment No. 220.Performance standards for the assessment of proposed similar modified in vitro reconstructed human epidermis (RhE) test methods for skin irritation testing as described in TG 4